

Minimum Complexity Pursuit: Stability Analysis

Shirin Jalali

Center for Mathematics of Information
California Institute of Technology
Pasadena, California, 91125

Arian Maleki

Department of Electrical Engineering
Rice University
Houston, Texas, 77005

Richard Baraniuk

Department of Electrical Engineering
Rice University
Houston, Texas, 77005

Abstract—A host of problems involve the recovery of structured signals from a dimensionality reduced representation such as a random projection; examples include sparse signals (compressive sensing) and low-rank matrices (matrix completion). Given the wide range of different recovery algorithms developed to date, it is natural to ask whether there exist “universal” algorithms for recovering “structured” signals from their linear projections. We recently answered this question in the affirmative in the noise-free setting. In this paper, we extend our results to the case of noisy measurements.

I. INTRODUCTION

Data compressors are ubiquitous in the digital world. They are built based on the premise that text, images, videos, etc. all are highly structured objects, and hence exploiting those structures can dramatically reduce the number of bits required for their storage. In recent years, a parallel trend has been developing for sampling analog signals. There too, the idea is that many analog signals of interest have some kind of structure that enables considerably lowering the sampling rate from the Shannon-Nyquist rate.

The first structure that was extensively studied in this context is *sparsity*. It has been observed that many natural signals have sparse representations in some domain. The term compressed sensing (CS) refers to the process of undersampling a high-dimensional sparse signal through linear measurements and recovering it from those measurements using efficient algorithms [1], [2]. Low-rankness [3], model-based compressed sensing [4]–[8], and finite rate of innovation [9] are examples of some other structures that have already been explored in the literature.

While in the original source coding problem introduced by Shannon [10], the assumption was that the source distribution is known both to the encoder and to the decoder, and hence is used in the code design, it was later shown that this information is not essential. In fact, *universal* compression algorithms are able to code stationary ergodic processes at their entropy rates, without knowing the source distribution [11]. In other words, there exists a family of compression codes that are able to code any stationary ergodic process at its entropy rate asymptotically [11]. The same result holds for universal lossy compression.

One can ask similar questions for the problem of undersampling “structured” signals: How to define the class of “structured” signals? Are there sampling and recovery algorithms for the reconstruction of “structured” signals from

their linear measurements without having the knowledge of the underlying structure? Does this ignorance incur a cost in the sampling rate?

In algorithmic information theory, *Kolmogorov complexity*, introduced by Solomonoff [12], Kolmogorov [13], and Chaitin [14], defines a universal notion of complexity for finite-alphabet sequences. Given a finite-alphabet sequence x , the Kolmogorov complexity of x , $K(x)$, is defined as the length of the shortest computer program that prints x and halts. In [15], extending the notion of Kolmogorov complexity to real-valued signals¹ by their proper quantization, we addressed some of the above questions. We introduced the minimum complexity pursuit (MCP) algorithm for recovering “structured” signals from their linear measurements. We showed that finding the “simplest” solution satisfying the linear measurements recovers the signal using many fewer measurements than its ambient dimension.

In this paper, we extend the results of [15] to the case where the measurements are noisy. We first propose an updated version of MCP that takes into account that the measurements are a linear transformation of the signal plus Gaussian noise. We prove that the proposed algorithm is stable with respect to the noise and derive bounds on its reconstruction error in terms of the sampling rate and the variance of the noise.

The organization of this paper is as follows. Section II defines the notation used throughout the paper. Section II-B defines Kolmogorov information dimension of a real-valued signal. Section III formally defines the MCP algorithm and reviews and extends some of the related results proved in [15]. Section IV considers the case of noisy measurements and proves that MCP is stable. Section V mentions some of the related work in the literature, and Section VI concludes the paper. Appendix A presents two useful lemmas used in the proofs.

II. DEFINITIONS

A. Notation

Calligraphic letters such as \mathcal{A} and \mathcal{B} denote sets. For a set \mathcal{A} , $|\mathcal{A}|$ and \mathcal{A}^c denote its size and its complement, respectively. For a sample space Ω and event set $\mathcal{A} \subseteq \Omega$,

¹These type of extensions are straightforward and have already been explored in [16].

$\mathbb{1}_{\mathcal{A}}$ denotes the indicator function of the event \mathcal{A} . Bold-faced lower case letters denote vectors. For a vector $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, its ℓ_p and ℓ_∞ norms are defined as $\|\mathbf{x}\|_p^p \triangleq \sum_{i=1}^n |x_i|^p$ and $\|\mathbf{x}\|_\infty \triangleq \max_i |x_i|$, respectively. For integer n , let I_n denote the $n \times n$ identity matrix.

For $x \in [0, 1]$, let $((x)_1, (x)_2, \dots), (x)_i \in \{0, 1\}$, denote the binary expansion of x , i.e., $x = \sum_{i=1}^\infty 2^{-i}(x)_i$. The m -bit approximation of x , $[x]_m$, is defined as $[x]_m \triangleq \sum_{i=1}^m 2^{-i}(x)_i$. Similarly, for a vector $(x_1, \dots, x_n) \in [0, 1]^n$, $[x^n]_m \triangleq ([x_1]_m, \dots, [x_n]_m)$.

B. Kolmogorov complexity

The Kolmogorov complexity of a finite-alphabet sequence \mathbf{x} with respect to a universal Turing machine \mathcal{U} is defined as the length of the shortest program on \mathcal{U} that prints \mathbf{x} and halts.² Let $K(\mathbf{x})$ denote the Kolmogorov complexity of binary string $\mathbf{x} \in \{0, 1\}^*$ $\triangleq \cup_{n \geq 1} \{0, 1\}^n$.

Definition 1: For real-valued $\mathbf{x} = (x_1, x_2, \dots, x_n) \in [0, 1]^n$, define the *Kolmogorov complexity of \mathbf{x} at resolution m* as

$$K^{[\cdot]_m}(\mathbf{x}) = K([x_1]_m, [x_2]_m, \dots, [x_n]_m).$$

Definition 2: The *Kolmogorov information dimension* of vector $(x_1, x_2, \dots, x_n) \in [0, 1]^n$ at resolution m is defined as

$$\kappa_{m,n} \triangleq \frac{K^{[\cdot]_m}(x_1, x_2, \dots, x_n)}{m}.$$

To clarify the above definition, we derive an upper bound for $\kappa_{m,n}$.

Lemma 1: For $(x_1, x_2, \dots) \in [0, 1]^\infty$ and any resolution sequence $\{m_n\}$, we have

$$\limsup_{n \rightarrow \infty} \frac{\kappa_{m,n}}{n} \leq 1.$$

Therefore, by Lemma 1, we call a signal *compressible*, if $\limsup_{n \rightarrow \infty} n^{-1} \kappa_{m,n} < 1$. As stated in the following proposition, Lemma 1's upper bound on $\kappa_{m,n}$ is achievable.

Proposition 1: Let $\{X_i\}_{i=1}^\infty \stackrel{iid}{\sim} \text{Unif}[0, 1]$. Then,

$$\frac{1}{mn} K^{[\cdot]_m}(X_1, X_2, \dots, X_n) \rightarrow 1$$

in probability.

III. MINIMUM COMPLEXITY PURSUIT

Consider the problem of reconstructing a vector $x_o^n \in [0, 1]^n$ from $d < n$ random linear measurements $y_o^d = Ax_o^n$. The MCP algorithm proposed in [15] reconstructs x_o^n from its linear measurements y_o^d by solving the following optimization problem:

$$\begin{aligned} \min_{x^n} \quad & K^{[\cdot]_m}(x_1, \dots, x_n) \\ \text{s.t.} \quad & Ax^n = y_o^d. \end{aligned} \quad (1)$$

Let the elements of $A \in \mathbb{R}^{d \times n}$, A_{ij} , be i.i.d. $\mathcal{N}(0, 1)$.³ Let $\hat{x}_o^n = \hat{x}_o^n(y_o^d, A)$ denote the output of (1) to inputs $y_o^d = Ax_o^n$ and A . Theorem 1 stated below is a generalization of Theorem 2 proved in [15].

²See Chapter 14 of [11] for the exact definition of a universal computer, and more details on the definition of Kolmogorov complexity.

³Note that in [15] we had assumed that $A_{i,j} \sim \mathcal{N}(0, 1/d)$.

Theorem 1: Assume that $x_o = (x_{o,1}, x_{o,2}, \dots) \in [0, 1]^\infty$. For integers m and n , let $\kappa_{m,n}$ denote the Kolmogorov information dimension of x_o^n at resolution m . Then, for any $\tau_n < 1$ and $t > 0$, we have

$$\begin{aligned} \mathbb{P} \left(\|x_o^n - \hat{x}_o^n\|_2 > \frac{\sqrt{nd^{-1} + t + 1} + 1}{\tau_n} \sqrt{n2^{-2m+2}} \right) \\ \leq 2^{2\kappa_{m,n}m} e^{\frac{d}{2}(1-\tau_n^2+2\log \tau_n)} + e^{-\frac{d}{2}t^2}. \end{aligned}$$

Theorem 1 can be proved following the steps used in the proof of Theorem 2 in [15]. To interpret this theorem, in the following we consider several interesting corollaries that follow from Theorem 1. Note that in all of the results, the logarithms are to the base of Euler's number e .

Corollary 1: Assume that $x_o = (x_{o,1}, x_{o,2}, \dots) \in [0, 1]^\infty$ and $m = m_n = \lceil \log n \rceil$. Let $\kappa_n \triangleq \kappa_{m_n,n}$. Then if $d_n = \lceil \kappa_n \log n \rceil$, for any $\epsilon > 0$, we have $\mathbb{P}(\|x_o^n - \hat{x}_o^n\|_2 > \epsilon) \rightarrow 0$, as $n \rightarrow \infty$.

Proof: For $m = m_n = \lceil \log n \rceil$ and $d_n = \lceil \kappa_n \log n \rceil$,

$$\begin{aligned} & (\sqrt{nd^{-1} + t + 1} + 1) \sqrt{n2^{-2m_n+2}} \\ & \leq 2 \left(\sqrt{\lceil \kappa_n \log n \rceil^{-1} + (t+1)n^{-1}} + \sqrt{n^{-1}} \right). \end{aligned} \quad (2)$$

Hence, fixing $t > 0$ and setting $\tau_n = \tau = 0.1$, for any $\epsilon > 0$ and large enough values of n we have

$$\frac{(\sqrt{nd^{-1} + t + 1} + 1) \sqrt{n2^{-2m+2}}}{\tau_n} \leq \epsilon.$$

Therefore, for n large enough,

$$\begin{aligned} \mathbb{P}(\|x_o^n - \hat{x}_o^n\|_2^2 > \epsilon) \\ \leq 2^{2\kappa_n \log n} e^{\frac{\kappa_n \log n}{2}(1-\tau^2+2\log \tau)} + e^{-\frac{d}{2}t^2} \\ \leq e^{1.4\kappa_n \log n} e^{-1.7\kappa_n \log n} + e^{-\frac{d}{2}t^2}, \end{aligned} \quad (3)$$

which shows that as $n \rightarrow \infty$, $\mathbb{P}(\|x_o^n - \hat{x}_o^n\|_2^2 > \epsilon) \rightarrow 0$. \square

According to Corollary 1, if the complexity of the signal is less than κ , then the number of linear measurements needed for asymptotically perfect recovery is, roughly speaking, at the order of $\kappa \log n$. In other words, the number of measurements is proportional to the complexity of the signal and only logarithmically proportional to its ambient dimension.

Corollary 2: Assume that $x_o = (x_{o,1}, x_{o,2}, \dots) \in [0, 1]^\infty$ and $m = m_n = \lceil \log n \rceil$. Let $\kappa_n \triangleq \kappa_{m_n,n}$. Then, if $d_n = \lceil 3\kappa_n \rceil$, we have

$$\mathbb{P} \left(\frac{1}{\sqrt{n}} \|x_o^n - \hat{x}_o^n\|_2 > \epsilon \right) \rightarrow 0,$$

as $n \rightarrow \infty$, for any $\epsilon > 0$.

Proof: Setting $\tau_n = n^{-0.5}$, $m = m_n = \lceil \log n \rceil$, and $d_n = \lceil 3\kappa_n \rceil$ in Theorem 1, it follows that

$$\begin{aligned} \mathbb{P} \left(\frac{1}{\sqrt{n}} \|x_o^n - \hat{x}_o^n\|_2 > 2\sqrt{d_n^{-1} + (t+1)n^{-1}} + 2\sqrt{n^{-1}} \right) \\ \leq 2^{2\kappa_n \log n} e^{1.5\kappa_n(1-n^{-1}-\log n)} + e^{-\frac{d}{2}t^2} \\ = e^{-(1.5-2\log 2)\kappa_n \log n + \kappa_n(1.5-1.5n^{-1})} + e^{-\frac{d}{2}t^2}. \end{aligned}$$

Since $1.5 - 2 \log 2 > 0$, for any $\epsilon > 0$ and n large enough, we have

$$2\sqrt{d_n^{-1} + (t+1)n^{-1}} + 2\sqrt{n^{-1}} < \epsilon.$$

It follows that $P(\frac{1}{\sqrt{n}}\|x_o^n - \hat{x}_o^n\|_2 > \epsilon) \rightarrow 0$, as $n \rightarrow \infty$. \square

In other words, if we are interested in the normalized mean square error, or per element squared distance, then $3\kappa_n$ measurements are sufficient.

IV. STABILITY ANALYSIS OF MCP

In the previous section we considered the case where the signal is exactly of low complexity and the measurements are also noise-free. In this section, we extend the results to noisy measurements, where $y_o^d = Ax_o^n + w^d$, with $w^d \sim \mathcal{N}(0, \sigma^2 I_d)$. Assuming that the complexity of the signal is known at the reconstruction stage, we consider the following reconstruction algorithm:

$$\begin{aligned} \arg \min_{x^n} \quad & \|Ax^n - y_o^d\|_2^2, \\ \text{s.t.} \quad & K^{[\cdot]^m}(x^n) \leq \kappa_{m,n}m. \end{aligned} \quad (4)$$

Note that $\kappa_{m,n}m$ is an upper bound on the Kolmogorov complexity of x_o at resolution m . The major issue of this section is to calculate the number of measurements required for robust recovery in noise.

Theorem 2: Assume that $x_o = (x_{o,1}, x_{o,2}, \dots) \in [0, 1]^\infty$. For integers m and n , let $\kappa_{m,n}$ denote the information dimension of x_o^n at resolution m . If $m = m_n = \lceil \log n \rceil$ and $d = 8r\kappa_{m,n}m$, where $r > 1$, then for any $\epsilon > 0$, we have

$$P\left(\|x_o^n - \hat{x}_o^n\|_2^2 > \frac{(2\kappa_{m,n}m)\sigma^2}{\rho d}\right) \rightarrow 0, \quad (5)$$

as $n \rightarrow \infty$, where $\rho \triangleq (1 - \sqrt{r^{-1}})^2/2$.

According Theorem 2, as long as $d > 8r\kappa_n \log n$ the algorithm is stable in the sense that the reconstruction error is proportional to the variance of the input noise. By increasing the number of measurements one may reduce the reconstruction error.

Proof: Since by definition $K^{[\cdot]^m}(x_o^n) = \kappa_{m,n}m_n$, x_o^n is also a feasible point in (4). But, by assumption, \hat{x}_o^n is the solution of (4). Therefore,

$$\begin{aligned} \|A\hat{x}_o^n - y_o^d\|_2^2 &\leq \|Ax_o^n - y_o^d\|_2^2 \\ &= \|Ax_o^n - Ax_o^n - w^d\|_2^2 = \|w^d\|_2^2. \end{aligned} \quad (6)$$

Expanding $\|A\hat{x}_o^n - y_o^d\|_2^2 = \|A\hat{x}_o^n - Ax_o^n - w^d\|_2^2$ in (6), it follows that

$$\|A(\hat{x}_o^n - x_o^n)\|_2^2 + \|w^d\|_2^2 - 2(w^d)^T A(\hat{x}_o^n - x_o^n) \leq \|w^d\|_2^2. \quad (7)$$

Canceling $\|w^d\|_2^2$ from both sides of (7), we obtain

$$\begin{aligned} \|A(\hat{x}_o^n - x_o^n)\|_2^2 &\leq 2(w^d)^T A(\hat{x}_o^n - x_o^n) \\ &\leq 2|(w^d)^T A(\hat{x}_o^n - x_o^n)|. \end{aligned}$$

Let $e_m^n \triangleq x_o^n - [x_o^n]_m$ and $\hat{e}_m^n \triangleq \hat{x}_o^n - [\hat{x}_o^n]_m$ denote the quantization errors of the original and the reconstructed signals, respectively. Using these definitions, and the Cauchy-Schwartz inequality, we find a lower bound for $\|A(\hat{x}_o^n - x_o^n)\|_2^2$ as

$$\begin{aligned} \|A(\hat{x}_o^n - x_o^n)\|_2^2 &= \|A([\hat{x}_o^n]_m + \hat{e}_m^n - [x_o^n]_m - e_m^n)\|_2^2 \\ &= \|A([\hat{x}_o^n]_m - [x_o^n]_m) + A(\hat{e}_m^n - e_m^n)\|_2^2 \\ &\geq \|A([\hat{x}_o^n]_m - [x_o^n]_m)\|_2^2 \\ &\quad - 2|(\hat{e}_m^n - e_m^n)^T A^T A([\hat{x}_o^n]_m - [x_o^n]_m)| \\ &\geq \|A([\hat{x}_o^n]_m - [x_o^n]_m)\|_2^2 \\ &\quad - 2\|A(\hat{e}_m^n - e_m^n)\|_2\|A([\hat{x}_o^n]_m - [x_o^n]_m)\|_2. \end{aligned} \quad (8)$$

On the other hand, again using our definitions plus the Cauchy-Schwartz inequality, we find an upper bound on $|(w^d)^T A(\hat{x}_o^n - x_o^n)|$ as

$$\begin{aligned} |(w^d)^T A(\hat{x}_o^n - x_o^n)| &= |([\hat{x}_o^n]_m - [x_o^n]_m + \hat{e}_m^n - e_m^n)^T A^T w^d| \\ &\leq |([\hat{x}_o^n]_m - [x_o^n]_m)^T A^T w^d| + |(\hat{e}_m^n - e_m^n)^T A^T w^d| \\ &\leq |([\hat{x}_o^n]_m - [x_o^n]_m)^T A^T w^d| + \|\hat{e}_m^n - e_m^n\|_2\|A^T w^d\|_2. \end{aligned} \quad (9)$$

For any $x \in [0, 1]$, $0 \leq x - [x]_m < 2^{-m}$. Therefore,

$$\|\hat{e}_m^n - e_m^n\|_2 \leq \sqrt{n}2^{-2m+2}. \quad (10)$$

Let set \mathcal{S} be the set of all vectors of length n that can be written as the difference of two vectors with complexity less than $\kappa_{m,n}m$; that is,

$$\mathcal{S} = \{h_1^n - h_2^n : K(h_1^n) \leq \kappa_{m,n}m, K(h_2^n) \leq \kappa_{m,n}m\}.$$

Note that $|\mathcal{S}| \leq 2^{2\kappa_{m,n}m}$. Define the event \mathcal{E}_1 as

$$\mathcal{E}_1 \triangleq \{\forall h^n \in \mathcal{S} : \|(w^d)^T Ah^n\| \leq t_1\|h^n\|_2\}.$$

For any fixed h^n , Ah^n is an i.i.d. zero-mean Gaussian vector of length d and variance $\|h^n\|_2^2$. Assuming that $\|h^n\|_2 = 1$ and applying Lemma 3, we obtain

$$\begin{aligned} P(|(w^d)^T Ah^n| \geq t_1) &= P(\|w^d\|_2 G \geq t_1) \\ &= P(\|w^d\|_2 G \geq t_1, \|w^d\|_2 \geq \sqrt{d}\sigma(1+t_2)) \\ &\quad + P(\|w^d\|_2 G \geq t_1, \|w^d\|_2 < \sqrt{d}\sigma(1+t_2)) \\ &\leq P(\|w^d\|_2 \geq \sqrt{d}\sigma(1+t_2)) \\ &\quad + P(G \geq t_1(\sqrt{d}\sigma(1+t_2))^{-1}) \\ &\leq e^{-dt_2^2/2} + e^{-\frac{t_1^2}{2\sigma^2 d(1+t_2)}}. \end{aligned} \quad (11)$$

Hence, by the union bound and the fact that $|\mathcal{S}| \leq 2^{2\kappa_{m,n}m}$ [11], we have

$$P(\mathcal{E}_1^c) \leq 2^{2\kappa_{m,n}m} \left(e^{-dt_2^2/2} + e^{-\frac{t_1^2}{2\sigma^2 d(1+t_2)}} \right). \quad (12)$$

Note that

$$\|A(\hat{e}_m^n - e_m^n)\|_2 \leq \sigma_{\max}(A)\|\hat{e}_m^n - e_m^n\|_2.$$

For $t_3 > 0$, define the event \mathcal{E}_2 as

$$\mathcal{E}_2^{(n)} \triangleq \left\{ \sigma_{\max}(A) < (1+t_3)\sqrt{d} + \sqrt{n} \right\}.$$

It can be proved that [17]

$$\mathbb{P} \left(\mathcal{E}_2^{(n),c} \right) \leq e^{-dt_3^2/2}. \quad (13)$$

But if $\sigma_{\max}(A) < (1+t_3)\sqrt{d} + \sqrt{n}$, then from (10)

$$\|A(\hat{e}_m^n - e_m^n)\|_2 \leq \left(1 + (1+t_3)\sqrt{\frac{d}{n}} \right) 2^{-m+1}n.$$

Define the event $\mathcal{E}_3^{(n)}$ as $\mathcal{E}_3^{(n)} \triangleq \{\forall h^n \in \mathcal{S} : \|Ah^n\|_2^2 > (1-t_4)d\|h^n\|_2^2\}$. By the union bound and Lemma 2, it follows that

$$\mathbb{P} \left(\mathcal{E}_3^{(n),c} \right) \leq 2^{2\kappa_{m,n}m} e^{\frac{d}{2}(t_4 + \log(1-t_4))}. \quad (14)$$

Define the event $\mathcal{E}_4^{(n)}$ as

$$\mathcal{E}_4^{(n)} \triangleq \{\forall h^n \in \mathcal{S} : \|Ah^n\|_2^2 < (1+t_5)d\|h^n\|_2^2\},$$

Again by the union bound and Lemma 2, it follows that

$$\mathbb{P} \left(\mathcal{E}_4^{(n),c} \right) \leq 2^{2\kappa_{m,n}m} e^{-\frac{d}{2}(t_5 - \log(1-t_5))}. \quad (15)$$

Finally, for $t_6 > 0$, define

$$\mathcal{E}_5^{(n)} \triangleq \{\|A^T w^d\|_2^2 \leq nd(1+t_6)\}.$$

Given w^d , $A^T w^d$ is an n dimensional i.i.d. Gaussian random vector with variance $\|w^d\|_2^2$. Hence, by Lemma 2,

$$\mathbb{P} \left(\|A^T w^d\|_2^2 \geq n\gamma^2(1+t_7) \mid \|w^d\|_2^2 = \gamma^2 \right) \leq e^{-\frac{n}{2}(t_7 - \log(1+t_7))}.$$

On the other hand, again by Lemma 2,

$$\mathbb{P}(\|w^d\|_2^2 < d(1-t_8)) \leq e^{\frac{d}{2}(t_8 + \log(1-t_8))}.$$

Choosing $t_6, t_7, t_8 > 0$ such that $t_6 < t_7$ and $1+t_6 = (1-t_8)(1+t_7)$, it follows that

$$\begin{aligned} \mathbb{P}(\|A^T w^d\|_2^2 \geq nd(1+t_6)) &= \mathbb{P}(\|A^T w^d\|_2^2 \geq nd(1+t_6), \|w^d\|_2^2 > d(1-t_8)) \\ &\quad + \mathbb{P}(\|A^T w^d\|_2^2 \geq nd(1+t_6), \|w^d\|_2^2 < d(1-t_8)) \\ &\leq e^{-\frac{n}{2}(t_7 - \log(1+t_7))} + e^{\frac{d}{2}(t_8 + \log(1-t_8))}. \end{aligned} \quad (16)$$

Combining (8) and (9) and the upper and lower bounds derived for the corresponding terms of (8) and (9), and choosing $t_1 = 2\sigma\sqrt{d(1+t_2)(2\kappa_{m,n}m)}$, with probability $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \mathcal{E}_4 \cap \mathcal{E}_5)$, the following inequality holds:

$$\begin{aligned} &(1-t_4)\sqrt{d}\|\Delta_m\|_2^2 \\ &- 2 \left(\sqrt{1+t_5}2^{-m+1}\sqrt{n}((1+t_3)\sqrt{d} + \sqrt{n}) \right) \|\Delta_m\|_2 \\ &- 2(\sigma\sqrt{1+t_2}\sqrt{2\kappa_{m,n}m}) \|\Delta_m\|_2 \\ &- 2^{-m+1}\sqrt{1+t_6}n \leq 0. \end{aligned} \quad (17)$$

Inequality (17) involves a quadratic equation of $\|\Delta_m\|_2$. Finding the roots of this quadratic equation, using $\sqrt{1+x} \leq 1+x/2$, and replacing $m = \lceil \log n \rceil$, we obtain

$$\begin{aligned} \|\Delta_m\|_2 &\leq \sigma\gamma_3\sqrt{(2\kappa_{m,n}\log n)d^{-1}} \\ &\quad + (\gamma_1\sqrt{n^{-1}} + \gamma_2\sqrt{d^{-1}}) + \sqrt{d^{-1}}\gamma_4, \end{aligned} \quad (18)$$

where $\gamma_1 = \sqrt{1+t_5}(1+t_3)(1-t_4)^{-1}$, $\gamma_2 = \sqrt{1+t_5}(1-t_4)^{-1}$, $\gamma_3 = \sqrt{1+t_2}(1-t_4)^{-1}$ and $\gamma_4 = \sqrt{1+t_6}(1-t_4)^{-1}$. On the other hand, by the union bound,

$$\begin{aligned} \mathbb{P}((\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \mathcal{E}_4 \cap \mathcal{E}_5)^c) &= \mathbb{P}(\mathcal{E}_1^c \cup \mathcal{E}_2^c \cup \mathcal{E}_3^c \cup \mathcal{E}_4^c \cup \mathcal{E}_5^c) \\ &\leq \mathbb{P}(\mathcal{E}_1^c) + \mathbb{P}(\mathcal{E}_2^c) + \mathbb{P}(\mathcal{E}_3^c) + \mathbb{P}(\mathcal{E}_4^c) + \mathbb{P}(\mathcal{E}_5^c). \end{aligned} \quad (19)$$

Given $d = 8r\kappa_{m,n}m$, choosing $t_2 = t_4 = 1/\sqrt{r}$ and fixing $t_1, t_3, t_5, \dots, t_8$ at appropriate fixed small numbers, (12), (13), (14), (15) and (16) guarantee that (19) goes to zero, as $n \rightarrow \infty$. Moreover, for chosen parameters, $\gamma_3 < \sqrt{2}/(1-\sqrt{r^{-1}})$. Finally, for any $\epsilon > 0$, for n large enough, $(\gamma_1\sqrt{n^{-1}} + \gamma_2\sqrt{d^{-1}}) + \sqrt{d^{-1}}\gamma_4 < \epsilon$. This concludes the proof. \square

V. RELATED WORK

The MCP algorithm proposed in [15] is mainly inspired by [18] and [19]. Consider the universal denoising problem where θ is corrupted by additive white Gaussian noise as $X^n = \theta + Z^n$. The denoiser's goal is to recover θ from the noisy observation X^n . The *minimum Kolmogorov complexity estimation* (MKCE) approach proposed in [18] suggests a denoiser that looks for the sequence $\hat{\theta}$ with minimum Kolmogorov complexity among all the vectors that are within some distance of the observation vector X^n . [18] shows that if θ_i are i.i.d., then under certain conditions, the average marginal conditional distribution of $\hat{\theta}_i$ given X_i tends to the actual posterior distribution of θ_1 given X_1 .

In [18], the authors consider the problem of recovering a low-complexity sequence from its linear measurements. Let $\mathcal{S}(k_0) \triangleq \{x^n \in [0,1]^n : K(x^n) \leq k_0\}$. Consider measuring $x_o^n \in \mathcal{S}(k_0)$ using a $d \times n$ binary matrix A . Let $y_o^d = Ax_o^n$. To recover x_o^n from measurements y_o^d , [18] suggests finding \hat{x}^n as $\hat{x}^n(y_o^d, A) \triangleq \arg \min_{x^n: y_o^d = Ax^n} K(x^n)$, and proves that if $d \geq 2k_0$, then this algorithm is able to find x_o^n with high probability. Clearly assuming that a real-valued sequence has a low complexity is very restrictive, and hence $\mathcal{S}(k_0)$ does not include any of the classes that has been studied in CS literature. For instance most of the one sparse signals have infinite Kolmogorov complexity, and hence the result of [18] does not imply useful information.

In a recent and independent work, [20] and [21] consider a scheme similar to MCP. For a stationary and ergodic source, they propose an algorithm to approximate MCP. While the empirical results are promising, no theoretical guarantees are provided on either the performance of MCP or their final algorithm.

The notion of sparsity has already been generalized in the literature in several different directions [3], [4], [9], [22]. More recently, [22] introduced the class of simple functions and atomic norm as a framework that unifies some

of the above observations and extends them to some other signal classes. While all these models can be considered as subclasses of the general model considered in this paper, it is worth noting that even though the recovery approach proposed here is universal, given the incomputibility of Kolmogorov complexity, it is not useful for practical purposes. Finding practical algorithms with provable performance guarantees is left for future research.

In this paper, we have focused on deterministic signal models. For the case of random signals, [23] considers the problem of recovering a memoryless process from its under-sampled linear measurements and establishes a connection between the required number of measurements and the Renyi entropy of the source. Also, our work is in the same spirit as the minimum entropy decoder proposed by Csiszar in [24], which is a universal decoder, for reconstructing an i.i.d. signal from its linear measurements.

VI. CONCLUSION

In this paper, we have considered the problem of recovering structured signals from their random linear measurements. We have investigated the minimum complexity pursuit (MCP) scheme. Our results confirm that if the Kolmogorov complexity of the signal is upper bounded by κ , then MCP recovers the signal accurately from $O(\kappa \log n)$ random linear measurements, which is much smaller than the ambient dimension. In this paper, we have specifically proved that MCP is stable, such that the ℓ_2 -norm of the reconstruction error is proportional to the standard deviation of the noise.

APPENDIX A

The following two lemmas are frequently used in our proofs.

Lemma 2 (χ -square concentration): Fix $\tau > 0$, and let $Z_i \sim \mathcal{N}(0, 1)$, $i = 1, 2, \dots, d$. Then, $P(\sum_{i=1}^d Z_i^2 < d(1 - \tau)) \leq e^{-\frac{d}{2}(\tau + \log(1 - \tau))}$, and

$$P(\sum_{i=1}^d Z_i^2 > d(1 + \tau)) \leq e^{-\frac{d}{2}(\tau - \log(1 + \tau))}.$$

The proof of Lemma 2 is presented in [15].

Lemma 3: Let X^n and Y^n denote two independent Gaussian random vectors with i.i.d. elements. Further assume that for $i = 1, \dots, n$, $X_i \sim \mathcal{N}(0, 1)$ and $Y_i \sim \mathcal{N}(0, 1)$. Then the distribution of $(X^n)^T Y^n = \sum_{i=1}^n X_i Y_i$ is the same as the distribution of $\|X^n\|_2 G$, where $G \sim \mathcal{N}(0, 1)$ is independent of $\|X^n\|_2$.

Proof: We need to show that $(X^n)^T Y^n / \|X^n\|_2$ is distributed as $\mathcal{N}(0, 1)$ and is independent of $\|X^n\|_2$. To prove the first claim, note that

$$\frac{(X^n)^T Y^n}{\|X^n\|_2} = \sum_{i=1}^n \frac{X_i}{\|X^n\|_2} Y_i. \quad (\text{A-1})$$

On the other hand, given $X^n / \|X^n\|_2 = a^n$,

$$\sum_{i=1}^n \frac{X_i}{\|X^n\|_2} Y_i \sim \mathcal{N}(0, 1),$$

because $\sum_{i=1}^n a_i^2 = 1$. Therefore, since the distribution of $\sum_{i=1}^n \frac{X_i}{\|X^n\|_2} Y_i$ given $X^n / \|X^n\|_2 = a^n$ is independent of a^n ,

$$\sum_{i=1}^n \frac{X_i}{\|X^n\|_2} Y_i \sim \mathcal{N}(0, 1).$$

To prove the independence, note that $X^n / \|X^n\|_2$ and Y^n are both independent of $\|X^n\|_2$. □

REFERENCES

- [1] D. L. Donoho. Compressed sensing. *IEEE Trans. Info. Theory*, 52(4):489–509, April 2006.
- [2] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Info. Theory*, 52(2):489–509, February 2006.
- [3] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, April 2010.
- [4] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. Info. Theory*, 56(4):1982–2001, April 2010.
- [5] Y. C. Eldar, P. Kuppinger, and H. Bolcskei. Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Trans. on Sig. Proc.*, 58(6):3042–3054, 2010.
- [6] M. Stojnic, F. Parvaresh, and B. Hassibi. On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Trans. on Sig. Proc.*, 57(8):3075–3085, 2009.
- [7] M. Stojnic. Block-length dependent thresholds in block-sparse compressed sensing. *Arxiv preprint arXiv:0907.3679*, 2009.
- [8] D. Malioutov, M. Cetin, and A.S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. on Sig. Proc.*, 53(8):3010–3022, August 2005.
- [9] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Trans. on Sig. Proc.*, 50(6):1417–1428, June 2002.
- [10] C. E. Shannon. A mathematical theory of communication: Parts I and II. *Bell Syst. Tech. J.*, 27:379–423 and 623–656, 1948.
- [11] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley, New York, 2nd edition, 2006.
- [12] R. J. Solomonoff. A formal theory of inductive inference. *Inform. Contr.*, 7:224–254, 1964.
- [13] A. N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Trans. Info. Theory*, 14:662–664, 1968.
- [14] G. J. Chaitin. On the length of program for computing binary sequences. *J. Assoc. Comput. Mach.*, 13:547–569, 1966.
- [15] S. Jalali and A. Maleki. Minimum complexity pursuit. In *Proc. 49th Allerton Conference on Communication, Control, and Computation*, Monticello, IL, Sep. 2011.
- [16] L. Staiger. The Kolmogorov complexity of real numbers. *Theoretical Computer Science*, 284(2):455–466, 2002.
- [17] E. Candès, J. Romberg, and T. Tao. Decoding by linear programming. *IEEE Trans. Info. Theory*, 51(12):4203–4215, Dec. 2005.
- [18] D. L. Donoho. Kolmogorov sampler. *Preprint*, 2002.
- [19] D. L. Donoho, H. Kakavand, and J. Mammen. The simplest solution to an underdetermined system of linear equations. In *Proc. IEEE Int. Symp. Info. Theory*, July 2006.
- [20] D. Baron and M. Duarte. Universal MAP estimation in compressed sensing. In *Proc. of 49th Allerton Conference on Communication, Control, and Computation*, Monticello, IL, Sep. 2011.
- [21] D. Baron and M. Duarte. Signal recovery in compressed sensing via universal priors. *Arxiv preprint arXiv:1204.2611*, 2012.
- [22] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. *Preprint*, 2010.
- [23] Y. Wu and S. Verdú. Renyi information dimension: Fundamental limits of almost lossless analog compression. *IEEE Trans. Info. Theory*, 56(8):3721–3748, Aug. 2010.
- [24] I. Csiszar. Linear codes for sources and source networks: Error exponents, universal coding. *IEEE Trans. Info. Theory*, 28:585–592, 1982.